Race and Subprime Loan Pricing *

Andra C. Ghent, Rubén Hernández-Murillo, and Michael T. Owyang†

This draft: February 2011.
Preliminary. Do not quote.

Abstract

In this paper we investigate whether race and ethnicity influenced subprime loan pricing during 2005, the peak of the subprime mortgage expansion. We combine loan-level data on the performance of non-prime securitized mortgages with individual- and neighborhood-level data on racial and ethnic characteristics for metropolitan areas in California and Florida.

Using a model of rate determination that accounts for predicted loan performance, we evaluate the presence of disparate impact and disparate treatment from race and ethnicity on rate-setting behavior across the most popular subprime mortgage products. In contrast with previous studies of the subprime market, we find evidence of adverse pricing effects for black and Hispanic borrowers.

Keywords: Fair Housing Act; Subprime Mortgages; Loan Performance; Discrimination.
JEL Codes: G21, J15, R23, C11

*The views expressed herein are those of the authors and do not reflect the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†Ghent: Zicklin School of Business, Baruch College / CUNY; phone 646-660-6929; email andra.ghent@baruch.cuny.edu. Hernández-Murillo: Research Division, Federal Reserve Bank of St. Louis; phone 314-444-8588; email: ruben.hernandez@stls.frb.org; Owyang: Research Division, Federal Reserve Bank of St. Louis; phone 314-444-8558; email owyang@stls.frb.org.

1 Introduction

A long literature examines the role of income and race on consumer lending. Most of this literature focuses on whether racial minorities are denied credit more frequently than white households with similar observable credit characteristics and on whether lenders deny credit to residents of neighborhoods with a high proportion of minorities (a practice called *redlining*).

Research on mortgages that originated prior to 1995, when mortgages were usually underwritten manually, found strong evidence that lenders were denying credit more frequently to black households than to white households with similar observable characteristics.[1] Financial and technological innovation in underwriting processes have made risk-based pricing of credit, rather than mere credit allocation, a more relevant issue in recent years. This is especially true in the subprime market where lenders were much less likely to sell the loan to government-sponsored enterprises (GSEs), such as Fannie Mae and Freddie Mac, and were thus less constrained by firm cutoffs on variables such as loan-to-value ratios, loan size, and credit scores. In a world where lenders cope with credit risk by rationing credit, discrimination and redlining manifest themselves primarily in loan denials. In contrast, when borrowers choose amongst several different sets of loan terms, each with a different price, minorities may be more able to obtain credit but may pay a higher price for it. Indeed, and perhaps in response to more stringent allocation constraints in prime mortgage markets, a disproportionate share of subprime loans went to black and Hispanic households (Mayer and Pence, 2008).

In this paper, we examine the influence of race and ethnicity on subprime loan pricing during 2005. Our study is most closely related to that of Haughwout, Mayer, and Tracy (2009). They matched data on loan pricing and risk measures with data on borrowers' race and racial composition of neighborhoods to analyze so-called 2/28 mortgages during August

[1] The seminal study is Munnell, Browne, McEneaney, and Tootell (1996). Ross and Yinger (2002) provide a comprehensive overview and analysis of the literature surrounding that study; see also Ladd (1998).

of 2005 for the entire United States.

Although we use a similar procedure to obtain the matched loan data, our analysis differs from theirs in many respects. First, we focus our analysis on California and Florida, two of the states with the highest incidence of subprime mortgages, and we extend the analysis to all of 2005 and across 8 different mortgage product categories. More importantly, we evaluate the presence of discrimination in loan pricing by analyzing the effect of race and neighborhood characteristics both on (1) the lenders' assessment of borrowers' risk profiles in an actuarial stage and (2) on the interest rate determination in an underwriting stage. Haughwout et al. do not consider that lenders' forecasts of loan performance are used in the underwriting process and that these forecasts may be correlated with race. Also, while their approach only allows them to evaluate disparate treatment discrimination, our methodology allows us to distinguish between disparate treatment and more subtle forms of discrimination. For example, as suggested by Ross and Tootell (2004), lenders may require black and Hispanic borrowers to purchase private mortgage insurance when they would not require a white borrower with a similar risk profile to do so.

As in Haughwout et al. (2009), we find no evidence of adverse loan pricing from race and ethnicity in 2/28 mortgages—although we find that lower-income neighborhoods face higher interest rates for this mortgage product, which suggests redlining. In contrast, we find that race and ethnicity had an adverse effect on loan pricing in various other mortgage products, resulting in increases in interest rates for blacks and Hispanics ranging from 5 to 35 basis points. Specifically, we find that Hispanics face higher interest rates in 5 of the 8 mortgage categories we analyzed, while blacks face higher interest rates in 3 mortgage categories. We also find evidence of redlining in 7 mortgage categories. Finally, we find that for blacks and Hispanics, purchasing private mortgage insurance and facing prepayment penalties seem to be associated with obtaining lower interest rates in some mortgage categories.

Additional recent papers that examine the effect of race on consumer credit include Woodward (2008), Reid and Laderman (2009), Pope and Sydnor (2008), and Ravina (2008).

Woodward (2008) examines closing costs and finds that they are higher for minorities and households with less education. Reid and Laderman (2009) study the link between race and ethnicity and the likelihood of obtaining higher priced loans in California. Pope and Sydnor (2008) and Ravina (2008) analyze the peer-to-peer lending market and find evidence of higher loan pricing for black borrowers when compared with white borrowers with similar risk profiles.

In the next section we describe the data and summarize the matching algorithm. In section 3 we present the model of rate determination and describe the estimation methodology. We present our results in section 4 and provide concluding remarks in section 5.

2 Data

Our data are non-prime, securitized, first-lien mortgages originated in 2005 in California and Florida. We merge detailed data on the performance and terms of the loans from First American LoanPerformance (LP) with data on borrower income, borrower race, Census tract income, and Census tract racial composition obtained under the Home Mortgage Disclosure Act (HMDA). To match loans from LP with HMDA data, we use a matching algorithm similar to that of Haughwout, Mayer, and Tracy (2009).

2.1 Matching LP data with HMDA data

The matching procedure considers first-lien loans with the same purpose (purchase or refinance) and occupancy status (owner-occupied). LP associates each loan with a 5-digit ZIP code, while HMDA loans are associated with Census tracts. To match ZIP codes with Census tracts we used Census ZIP Code Tabulation Areas (ZCTAs).[2] We also use GIS software to establish Census tracts search areas associated with any given ZCTAs as follows: for each loan in LP we determined the smallest set of Census tracts that intersect with the associated

[2]ZCTAs are statistical entities developed by the Census for tabulating summary statistics from the 2000 Census for geographic areas that approximate the land area covered by each ZIP code.

ZCTA and we allowed for the union of the Census tracts in the intersection to extend over the geographic area defined by any given ZCTA.

Except for the use of ZCTAs, we followed Haughwout et al.'s matching algorithm very closely. The procedure entails 6 stages which use the originator's name, the loan amount, and the origination dates to obtain the matches. The names are provided by the lenders themselves in the HMDA data, but not in the LP data. As a result, lender names in LP have to be cleaned manually before the matching. Loan amounts are provided in dollars in LP, while they are provided in thousands in HMDA. Furthermore, HMDA allows lenders to round up loan amounts to the nearest thousand if the fraction equals or exceeds $500. The dates are matched to within 5 business days if the LP dates are not imputed or to the same month if they are.[3] A summary of the various stages is as follows:

- Stage 1 considers loans with matched originator names and uses the larger 4-digit ZCTA search areas. Loan amounts are matched allowing a difference of up to and including $1,000.
- Stage 2 ignores originator names and uses 4-digit ZCTA search areas, as in stage 1.
- Stage 3 again considers originator names, but uses the smaller 5-digit ZCTA search areas. Loan amounts are matched allowing a difference of up to but not including $1,000.
- Stage 4 is similar to stage 3 but ignores originator names.
- Stage 5 is similar to stage 1 but loan amounts are matched to within 2.5% of the LP amount.
- Stage 6 is similar to stage 2 but loan amounts are matched to within 2.5% of the LP amount.

[3]LP origination dates are considered to be imputed if they are exactly two months before the first payment date.

At the conclusion of each stage, only one-to-matches are kept and are removed from the data sets, while loans with multiple matches (either one LP loan to many HMDA loans, or many LP loans to one HMDA loan) are thrown back into the matching pool for the subsequent stages.

2.2 Summary statistics

Tables 1 through 6 contain summary statistics on the loans in our sample by race and product type. Table 1 summarizes the counts of mortgages by product and race that were matched. We consider three racial or ethnic categories: Hispanics, non-Hispanic blacks, and the remainder (non-Hispanic and non-blacks). We also consider the largest seven non-prime mortgage categories (which account for about 90 percent of all non-prime loans) and we included a category for the remainder. In addition to 2/28 adjustable rate mortgages (ARMs) (with a fixed interest rate for the first two years and a variable rate for the remaining 28 years), we also consider 3/27 ARMs (with a fixed rate for the first three years and a variable rate for the remaining 27 years), 10 year ARMs, 10 year fixed rate mortgages (FRMs), 5 year ARMs, 30 year ARMs, and 30 year FRMs. As can be gleaned from the table, except for 3/27 ARMs and 10 year ARMs, all other categories contain more loans than the 2/28 category.

We matched 283,180 purchase loans and 380,195 refinances. Hispanic borrowers obtained 102,230 purchase loans, almost 5 times the amount for black borrowers, and they obtained almost 98,000 refinancing loans, about 3 times the amount for black borrowers. The most popular products for home purchases across all race categories were 2/28 ARMs, 30 year ARMs, and 5 year AMRs. For refinances the most popular products also included 30 year FRMs. For comparison, note that Haughwout et al. (2009) matched only 2/28 ARMs using national data for August of 2005 for a total of about 75,000 loans.

Table 2 summarizes the proportion of loans by product and racial groups that (1) included prepayment penalties (Prepay), (2) required purchase of private mortgage insurance

Table 1: Mortgage Counts

	Purchases				Refinances				
Product	Hispanic	black	other	Total	Hispanic	black	other	Total	Sum
10yr ARM	6,987	1,042	18,437	26,466	2,373	602	9,961	12,936	39,402
10yr FRM	1,394	250	4,868	6,512	1,290	311	6,007	7,608	14,120
2/28 ARM	10,029	1,468	10,070	21,567	4,231	1,144	7,170	12,545	34,112
3/27 ARM	2,434	460	4,363	7,257	1,499	484	3,518	5,501	12,758
30yr ARM	34,702	9,356	56,625	100,683	46,830	17,707	119,404	183,941	284,624
30yr FRM	4,295	1,058	10,321	15,674	16,762	6,619	44,445	67,826	83,500
5yr ARM	29,542	4,934	41,464	75,940	13,349	3,985	29,629	46,963	122,903
Other	12,847	2,012	14,222	29,081	11,603	3,780	27,492	42,875	71,956
Total	102,230	20,580	160,370	283,180	97,937	34,632	247,626	380,195	663,375

(PMI), and (3) required full documentation (Full Doc). Unconditionally, black and Hispanic borrowers face prepayment penalties more frequently than other borrowers in all product categories. The exception is that black borrowers face prepayment penalties at about the same frequency as other borrowers for 10 year FRM refinances. Also, both black and Hispanic borrowers tend to be required to obtain private mortgage insurance more often than other borrowers for most mortgage products. Finally, black borrowers are also required to provide full documentation slightly more often than Hispanics and other borrowers.

As tables 3 and 4 indicate, black and Hispanic borrowers tend to have lower FICO scores across most mortgage products (except that for 2/28s Hispanic borrowers show a slightly higher FICO score than other borrowers). Black and Hispanic borrowers also tend to have mortgages with higher loan-to-value (LTV) ratios, and higher debt-to-income (DTI) ratios. The variable *Good Credit* summarizes these differences; *Good Credit* takes a value of 1 if the borrower has a FICO score above the 50th percentile, the LTV is at or below the 50th percentile, and the DTI is at or below the 50th percentile. In summary, a smaller proportion of black and Hispanic borrowers exhibit good credit when compared with other borrowers both for purchases and for refinances.

Tables 5 and 6 summarize the loan amounts and contract interest rates. They also provide the average spread of the loan's annual percentage rate with a treasury security of comparable maturity, as provided to HMDA. Loan amounts for blacks and Hispanics are smaller than for other borrowers, and loan amounts for blacks are almost always smaller than for Hispanics. Loan amounts for purchases tend to be higher than for refinances. Contract rates and spreads are slightly lower on refinancing mortgages than on purchase mortgages. Black and Hispanic borrowers generally face higher contract rates than other borrowers; the difference in the rates that black and Hispanic borrowers pay relative to other borrowers is somewhat less pronounced in the spreads.

Table 2: Prepayment Penalties, Private Mortgage Insurance, and Full Documentation

		Purchases				Refinances			
Product	Race	N	Prepay	PMI	Full Doc	N	Prepay	PMI	Full Doc
10yr ARM	Hispanic	6,987	0.482	0.051	0.139	2,373	0.456	0.039	0.241
	black	1,042	0.427	0.076	0.243	602	0.445	0.055	0.287
	other	18,437	0.350	0.068	0.232	9,961	0.344	0.026	0.308
	Total	26,466	0.388	0.064	0.208	12,936	0.370	0.030	0.295
10yr FRM	Hispanic	1,394	0.311	0.072	0.220	1,290	0.350	0.031	0.367
	black	250	0.264	0.064	0.276	311	0.264	0.026	0.492
	other	4,868	0.259	0.052	0.328	6,007	0.274	0.016	0.437
	Total	6,512	0.271	0.056	0.303	7,608	0.286	0.019	0.427
2/28	Hispanic	10,029	0.945	0.094	0.352	4,231	0.972	0.129	0.507
	black	1,468	0.928	0.095	0.471	1,144	0.956	0.121	0.603
	other	10,070	0.906	0.098	0.405	7,170	0.949	0.125	0.587
	Total	21,567	0.926	0.096	0.385	12,545	0.957	0.126	0.562
3/27	Hispanic	2,434	0.684	0.090	0.414	1,499	0.844	0.103	0.548
	black	460	0.691	0.091	0.530	484	0.866	0.070	0.690
	other	4,363	0.506	0.056	0.424	3,518	0.737	0.076	0.603
	Total	7,257	0.578	0.069	0.427	5,501	0.777	0.083	0.596
30yr ARM	Hispanic	34,702	0.901	0.179	0.317	46,830	0.934	0.193	0.394
	black	9,356	0.920	0.209	0.446	17,707	0.946	0.217	0.520
	other	56,625	0.813	0.143	0.323	119,000	0.895	0.190	0.449
	Total	101,000	0.853	0.162	0.332	184,000	0.910	0.193	0.442
30yr FRM	Hispanic	4,295	0.655	0.206	0.389	16,762	0.851	0.183	0.582
	black	1,058	0.736	0.238	0.507	6,619	0.905	0.214	0.687
	other	10,321	0.479	0.146	0.420	44,445	0.778	0.188	0.651
	Total	15,674	0.545	0.168	0.417	67,826	0.808	0.189	0.637
5yr ARM	Hispanic	29,542	0.896	0.152	0.383	13,349	0.910	0.210	0.489
	black	4,934	0.879	0.129	0.521	3,985	0.913	0.194	0.607
	other	41,464	0.795	0.112	0.475	29,629	0.840	0.198	0.585
	Total	75,940	0.840	0.129	0.442	46,963	0.866	0.201	0.559
Other	Hispanic	12,847	0.911	0.073	0.210	11,603	0.915	0.133	0.397
	black	2,012	0.898	0.087	0.360	3,780	0.925	0.139	0.506
	other	14,222	0.767	0.083	0.275	27,492	0.863	0.124	0.459
	Total	29,081	0.840	0.079	0.252	42,875	0.883	0.128	0.446

Prepay, PMI, and Full Doc indicate the shares of mortgages in each category
with prepayment penalties, private mortgage insurance, and full documentation requirements.

Table 3: Borrower's Credit Characteristics. Purchases

			Good Credit	FICO		LTV		DTI	
Product	Race	N	Share	Mean	SD	Mean	SD	Mean	SD
10yr ARM	Hispanic	6,987	0.460	715.6	47.8	79.6	5.3	25.3	18.8
	black	1,042	0.430	709.8	44.0	80.0	5.7	26.5	18.5
	other	18,437	0.480	722.3	46.1	79.2	7.1	25.8	18.1
	Total	26,466	0.470	720.0	46.6	79.3	6.7	25.7	18.3
10yr FRM	Hispanic	1,394	0.600	718.2	46.3	78.5	8.3	14.7	19.4
	black	250	0.600	712.1	47.8	78.7	8.3	14.0	19.1
	other	4,868	0.640	725.9	46.6	76.7	10.2	14.1	19.1
	Total	6,512	0.630	723.7	46.7	77.1	9.8	14.2	19.2
2/28	Hispanic	10,029	0.170	668.6	45.7	81.3	4.7	32.8	18.5
	black	1,468	0.140	655.9	45.5	82.4	5.8	32.3	18.6
	other	10,070	0.160	665.0	47.8	81.4	4.9	32.6	18.7
	Total	21,567	0.160	666.0	46.8	81.4	4.9	32.7	18.6
3/27	Hispanic	2,434	0.310	677.2	55.1	81.8	5.7	19.1	20.6
	black	460	0.250	665.3	59.6	81.7	6.3	17.5	20.5
	other	4,363	0.370	686.6	58.7	80.9	6.2	17.9	20.3
	Total	7,257	0.340	682.1	57.9	81.2	6.0	18.3	20.4
30yr ARM	Hispanic	34,702	0.200	657.9	58.8	82.1	7.4	27.8	20.3
	black	9,356	0.130	635.3	59.4	84.0	7.8	28.1	20.3
	other	56,625	0.320	670.3	67.3	80.7	8.5	24.0	20.5
	Total	100,683	0.260	662.8	64.6	81.5	8.2	25.7	20.5
30yr FRM	Hispanic	4,295	0.350	683.7	60.2	78.9	12.5	21.1	21.1
	black	1,058	0.250	661.1	68.3	80.5	12.1	22.8	21.4
	other	10,321	0.470	696.1	66.0	76.6	13.9	16.7	20.3
	Total	15,674	0.420	690.4	65.3	77.5	13.5	18.3	20.7
5yr ARM	Hispanic	29,542	0.200	676.3	47.9	80.9	4.7	34.0	18.3
	black	4,934	0.160	662.0	49.1	81.5	5.4	34.5	18.6
	other	41,464	0.230	678.0	51.7	80.6	5.6	32.6	19.1
	Total	75,940	0.210	676.3	50.2	80.8	5.3	33.3	18.8
Other	Hispanic	12,847	0.190	665.3	51.6	80.5	5.0	32.4	19.3
	black	2,012	0.150	648.2	53.5	81.3	5.9	32.2	19.6
	other	14,222	0.300	679.6	61.7	79.7	7.3	28.4	19.8
	Total	29,081	0.240	671.2	57.6	80.2	6.3	30.4	19.7

The variable *Good Credit* takes a value of 1 if the borrower has a FICO score above the 50th percentile, Loan-to-Value ratio at or below the 50th percentile, and Debt-to-Income ratio at or below the 50th percentile.

Table 4: Borrower's Credit Characteristics. Refinances

			Good Credit	FICO		LTV		DTI	
Product	Race	N	Share	Mean	SD	Mean	SD	Mean	SD
10yr ARM	Hispanic	2,373	0.460	697.4	45.9	71.4	12.3	24.4	18.8
	black	602	0.410	695.2	49.0	72.8	11.9	25.4	18.7
	other	9,961	0.540	710.6	47.5	69.5	13.3	24.7	17.9
	Total	12,936	0.520	707.5	47.6	70.0	13.1	24.7	18.1
10yr FRM	Hispanic	1,290	0.590	700.1	48.4	65.7	14.6	13.9	18.8
	black	311	0.630	705.0	49.4	66.4	14.6	12.9	18.8
	other	6,007	0.680	715.5	50.2	64.5	15.5	13.1	18.3
	Total	7,608	0.660	712.5	50.2	64.8	15.3	13.3	18.4
2/28	Hispanic	4,231	0.070	640.1	43.2	80.8	11.4	32.6	17.7
	black	1,144	0.050	627.9	38.3	80.5	11.8	32.0	18.2
	other	7,170	0.060	632.6	41.7	80.6	11.6	31.1	18.7
	Total	12,545	0.060	634.7	42.1	80.7	11.5	31.7	18.4
3/27	Hispanic	1,499	0.180	644.6	51.5	77.2	12.4	17.9	20.4
	black	484	0.140	635.3	51.3	78.3	12.3	19.1	20.4
	other	3,518	0.220	646.9	56.6	76.8	12.4	15.5	19.9
	Total	5,501	0.200	645.2	54.9	77.0	12.4	16.5	20.1
30yr ARM	Hispanic	46,830	0.160	615.4	70.3	73.8	13.3	27.5	19.9
	black	17,707	0.090	594.9	64.4	75.5	12.9	28.8	19.9
	other	119,404	0.230	627.1	77.4	73.2	13.6	24.9	20.1
	Total	183,941	0.200	621.0	75.2	73.6	13.5	25.9	20.1
30yr FRM	Hispanic	16,762	0.210	641.0	62.7	67.2	15.9	23.4	21.1
	black	6,619	0.140	620.3	62.8	70.3	15.9	24.8	20.9
	other	44,445	0.260	647.8	68.5	68.6	16.4	21.6	20.7
	Total	67,826	0.240	643.4	67.0	68.4	16.3	22.3	20.9
5yr ARM	Hispanic	13,349	0.110	647.0	48.1	78.7	11.9	32.9	17.8
	black	3,985	0.090	638.1	45.0	79.7	11.5	32.5	18.2
	other	29,629	0.150	650.0	50.9	78.0	12.4	31.4	18.6
	Total	46,963	0.130	648.2	49.8	78.4	12.2	31.9	18.4
Other	Hispanic	11,603	0.190	635.4	65.3	71.6	15.6	29.2	19.3
	black	3,780	0.150	621.6	64.0	73.1	15.0	30.2	19.2
	other	27,492	0.290	652.8	73.0	70.9	15.8	27.5	19.0
	Total	42,875	0.250	645.3	71.0	71.3	15.7	28.2	19.1

The variable *Good Credit* takes a value of 1 if the borrower has a FICO score above the 50th percentile, Loan-to-Value ratio at or below the 50th percentile, and Debt-to-Income ratio at or below the 50th percentile.

Table 5: Loan Amount and Contract Rate. Purchases

			Loan Amount		Contract Rate		HMDA Spread	
Product	Race	N	Mean	SD	Mean	SD	Mean	SD
10yr ARM	Hispanic	6,987	351,587	157,989	6.17	0.64	4.61	0.87
	black	1,042	344,020	179,952	6.18	0.69	4.58	0.83
	other	18,437	415,350	242,788	6.01	0.65	4.55	1.02
	Total	26,466	395,708	223,189	6.06	0.65	4.57	0.94
10yr FRM	Hispanic	1,394	330,075	167,893	6.44	0.58	4.63	0.90
	black	250	329,845	180,702	6.51	0.60	4.41	1.03
	other	4,868	385,144	234,578	6.29	0.50	4.46	1.26
	Total	6,512	371,232	221,327	6.33	0.53	4.52	1.12
2/28	Hispanic	10,029	314,674	119,294	6.76	0.71	4.45	0.66
	black	1,468	299,742	130,789	6.85	0.80	4.49	0.79
	other	10,070	334,131	134,197	6.78	0.78	4.45	0.73
	Total	21,567	322,742	127,734	6.78	0.75	4.45	0.70
3/27	Hispanic	2,434	302,841	123,809	6.50	0.85	4.46	0.80
	black	460	283,932	150,799	6.66	0.90	4.61	0.81
	other	4,363	354,771	177,342	6.33	0.94	4.48	0.92
	Total	7,257	332,864	161,949	6.41	0.91	4.48	0.86
30yr ARM	Hispanic	34,702	279,095	164,854	6.59	1.82	4.71	0.89
	black	9,356	236,245	161,569	7.19	1.55	4.96	0.94
	other	56,625	369,974	270,337	6.06	2.20	4.79	0.99
	Total	100,683	326,225	235,560	6.35	2.05	4.78	0.94
30yr FRM	Hispanic	4,295	252,306	147,372	6.77	0.85	4.22	0.93
	black	1,058	207,940	165,064	7.21	1.10	4.33	1.06
	other	10,321	295,783	218,720	6.60	0.87	4.21	0.98
	Total	15,674	277,940	200,020	6.69	0.90	4.23	0.98
5yr ARM	Hispanic	29,542	320,809	132,203	6.63	0.76	4.54	0.78
	black	4,934	305,553	154,185	6.72	0.83	4.58	0.83
	other	41,464	351,149	176,499	6.49	0.80	4.44	0.84
	Total	75,940	336,383	160,086	6.56	0.79	4.49	0.82
Other	Hispanic	12,847	331,221	143,487	6.93	1.06	4.67	0.85
	black	2,012	316,266	163,512	7.16	1.12	4.85	0.92
	other	14,222	402,025	221,213	6.55	1.51	4.74	0.93
	Total	29,081	364,813	190,298	6.76	1.32	4.71	0.89

HMDA spread denotes the spread between the APR and the yield on a treasury security of comparable maturity if the loan is a *high cost* loan, defined as one for which the spread is 300 basis points or more.

Table 6: Loan Amount and Contract Rate. Refinances

			Loan Amount		Contract Rate		HMDA Spread	
Product	Race	N	Mean	SD	Mean	SD	Mean	SD
10yr ARM	Hispanic	2,373	366,210	197,168	6.04	0.66	4.60	0.96
	black	602	375,569	228,459	6.10	0.77	4.65	1.00
	other	9,961	478,403	300,767	5.86	0.75	4.61	1.02
	Total	12,936	453,036	285,249	5.90	0.74	4.61	1.00
10yr FRM	Hispanic	1,290	319,933	170,739	6.19	0.46	4.69	1.01
	black	311	319,736	174,209	6.21	0.46	4.88	1.02
	other	6,007	394,420	253,192	6.12	0.43	4.64	1.13
	Total	7,608	378,737	240,237	6.14	0.44	4.68	1.08
2/28	Hispanic	4,231	318,713	118,398	6.64	0.73	4.49	0.72
	black	1,144	315,450	126,163	6.68	0.78	4.49	0.82
	other	7,170	346,710	145,610	6.67	0.77	4.43	0.80
	Total	12,545	334,417	136,004	6.66	0.76	4.46	0.78
3/27	Hispanic	1,499	302,973	119,981	6.36	0.79	4.43	0.77
	black	484	292,605	139,361	6.40	0.79	4.49	0.83
	other	3,518	349,173	179,558	6.30	0.86	4.38	0.82
	Total	5,501	331,607	163,700	6.33	0.83	4.41	0.81
30yr ARM	Hispanic	46,830	271,409	144,078	6.55	2.05	4.87	1.00
	black	17,707	236,664	142,976	7.07	1.88	5.11	1.10
	other	119,404	328,301	235,372	6.34	2.25	5.02	1.25
	Total	183,941	304,996	210,486	6.46	2.17	4.99	1.17
30yr FRM	Hispanic	16,762	230,994	118,860	6.65	0.83	4.43	1.07
	black	6,619	194,692	118,838	7.03	1.02	4.44	1.15
	other	44,445	255,551	173,818	6.69	0.95	4.41	1.21
	Total	67,826	243,543	158,218	6.72	0.94	4.42	1.16
5yr ARM	Hispanic	13,349	320,576	127,922	6.60	0.77	4.58	0.83
	black	3,985	320,525	137,162	6.66	0.80	4.62	0.89
	other	29,629	361,895	180,343	6.53	0.82	4.50	0.88
	Total	46,963	346,640	164,815	6.56	0.80	4.54	0.87
Other	Hispanic	11,603	292,548	146,004	6.65	1.52	4.91	1.00
	black	3,780	279,258	156,582	6.88	1.52	5.04	2.13
	other	27,492	349,599	227,808	6.41	1.77	4.90	1.16
	Total	42,875	327,958	205,072	6.52	1.69	4.92	1.27

HMDA spread denotes the spread between the APR and the yield on a treasury security of comparable maturity if the loan is a *high cost* loan, defined as one for which the spread is 300 basis points or more.

3 A Model of Mortgage Rate Determination

In this section, we present a simple reduced-form model of mortgage rate determination which is derived from a test proposed in Ross and Yinger (2002, ch. 10). In the model, lenders charge a rate based on the expected performance of the loan. Loan performance is judged by the expected probability that it produces adverse outcomes—e.g., default or prepayment. Along the lines of Ladd (1998), who discusses various definitions of mortgage discrimination in light of the relevant mortgage laws, we allow for the possibility that lenders may vary the rate charged based on variables used to identify two broad classes of discrimination: *disparate treatment* and *disparate impact*. The former is manifest in rate changes directly associated with race variables. The latter occurs when policies that do not explicitly take race into account result in disparities among racial groups because race is correlated with other non-race variables that may be used in underwriting, even when they are not necessarily good predictors of loan performance. To this end, we allow loan performance to vary with race and other variables.

The advantage of this approach is that it enables us to detect both disparate impact and disparate treatment discrimination, both of which are illegal. In particular, if lenders wish to discriminate against a particular group, either because of *taste-based discrimination* (manifested in a direct effect of race and ethnicity on interest rates independent of the effect via loan performance) or because of *statistical discrimination* (manifested through the effect on predicted loan performance), lenders may change the weights of various loan characteristics in a pricing model to indirectly discriminate against minorities. Furthermore, by including Census tract characteristics, namely median family income and percent of minority population, we can also detect redlining.

The test proceeds as follows:

1. We draw a sample of loans for a particular mortgage product and estimate loan performance models (using default and prepayment as the adverse outcomes) using loan,

individual, and Census tract characteristics *including* the minority status of the borrower, the income of the Census tract, and the racial composition of the Census tract. We label this the *actuarial* stage.

2. We then draw a new sample of loans, and using the estimation outcomes from stage 1, we compute the *predicted* performance of the new sample of loans using loan and individual characteristics. In this step, we *omit* the minority status of the borrower, the Census tract income, and the racial composition of the Census tract.

3. Finally, we estimate a model with the loans from stage 2 using the actual interest rate as the dependent variable and the predicted probabilities of default and prepayment. We label this the *underwriting* stage.

3.1 Empirical Framework

To formalize, consider the following linear rate setting equation:

$$R_n = \beta_0 + \beta_p \widehat{\mathbf{P}}_n + \beta_z \mathbf{z}_n + \beta_x \mathbf{x}_n + e_n, \tag{1}$$

where R_n is the rate charged for loan n, $\widehat{\mathbf{P}}_n$ is a $(\pi \times 1)$ vector of measures of loan performance, $\mathbf{z}_n$ is a $(\kappa_z \times 1)$ vector of *impact* variables (non-race variables), and $e_n \sim N(0, \sigma^2)$. The $(\kappa_x \times 1)$ vector of *treatment* variables $\mathbf{x}_n$ includes a set of individual discrimination indicators (i.e., borrower race) and a set of redlining indicators (e.g., neighborhood racial composition).

In order to estimate (1), we require the vector of predicted loan performance measures, $\widehat{\mathbf{P}}_n$. Loan performance data typically consists of binary measures—i.e., does the loan default or gets prepaid within two years —which would not be available at the time the rate is set. Instead, we use the vector of expected loan performance, which is composed of the forecasted probability of loan default and the forecasted probability of prepayment. To construct these, we extract from the full sample of loans a subset of loans to use as an *actuarial* sample. From this sample, we estimate models of loan performance and use the resulting estimation to construct predicted performance for loans in a different *underwriting* sample on which we evaluate the presence of discrimination.

We partition the full set of loans into an M loan actuarial sample and an N loan underwriting sample. Let $\mathbf{P}_m$ represent the vector of π different performance measures for loan m from the actuarial sample. Let $\mathbf{q}_m$ represent the $(\kappa_q \times 1)$ vector of non-racial characteristics which affect loan performance (e.g., FICO score, loan-to-value ratio, etcetera), and let $\mathbf{w}_m$ represent the $(\kappa_w \times 1)$ vector of racial characteristics (black and Hispanic indicators) which may affect loan performance. For any m, the probability that $P_{im} = 1$— e.g., that loan m defaults—can be specified as a probit:

$$\Pr\left[P_{im} = 1\right] = \Phi\left(\alpha_{i0} + \alpha_{iq}\mathbf{q}_m + \alpha_{iw}\mathbf{w}_m\right), \tag{2}$$

where the link function, $\Phi\left(.\right)$, is the standard normal cdf and $\alpha_i = \left[\alpha_{i0}, \alpha_{iq}, \alpha_{iw}\right]$ are slope coefficients specific to the ith performance measure. From (2), the predicted probabilities for loans from the underwriting subsample are computed as

$$\widehat{P}_{in} = \Phi\left(\alpha_{i0} + \alpha_{iq}\mathbf{q}_n\right), \tag{3}$$

where, again, $\Phi\left(.\right)$ is the standard normal cdf. Note that the vector of race variables, w_m, are excluded from the calculation of the predicted loan performance measures. The use of these variables as predictors of loan performance are illegal; however, we must extract out their effect in the loan performance model in order to properly assess the effect of other measures.[4]

[4]We discuss below under what circumstances these treatment variables might be used in the predicted probabilities.

3.2 Identifying Types of Discrimination

As indicated above, we broadly classified three forms of discrimination: disparate treatment, disparate impact, and redlining. Disparate treatment discrimination will increase the rate charged to a minority borrower; redlining will increase the rate charged to individuals in a minority neighborhood. We differentiate disparate treatment discrimination from redlining by partitioning the treatment variable, $\mathbf{x}_n$, in the rate equation into individual and neighborhood subvectors, $\mathbf{x}_n^{ind}$ and $\mathbf{x}_n^{area}$, respectively. An increase in the rate attributable to elements of $\mathbf{x}_n^{ind}$ reflects disparate treatment discrimination while an increase in the rate attributable to elements of $\mathbf{x}_n^{area}$ reflects redlining. The use of variables that do not explicitly take race into account (included in the vector $\mathbf{z}_n$) and are not necessarily good predictors of performance might result in disparate impact on certain racial groups. We allow for the interactions of race and ethnicity indicators with impact variables to be included in the vector $\mathbf{x}_n$.

Discrimination may result from tasted-based discrimination (animosity or prejudice against minorities) or from statistical discrimination (the lender uses race or ethnicity to estimate the borrower's credit worthiness). To differentiate the two forms, the predicted loan performance used in underwriting (3) is rewritten to include the treatment variable, $\mathbf{w}_m$. In this case, discrimination causes a change in the loan's predicted performance through a difference in the probability of, say, default. To capture this, we can compute the predicted performance when race is included:

$$\widetilde{P}_{in} = \Phi\left(\alpha_{i0} + \alpha_{iq}\mathbf{q}_n + \alpha_{iw}\mathbf{w}_m\right). \tag{4}$$

and define the difference as $\Delta\widehat{P}_{in} = \widetilde{P}_{in} - \widehat{P}_{in}$. We can modify the rate equation to account for the change in expected loan performance:

$$R_n = \beta_0 + \beta_p\widehat{\mathbf{P}}_n + \beta_p\Delta\widehat{\mathbf{P}}_n + \beta_z\mathbf{z}_n + \beta_x\mathbf{x}_n + e_n, \tag{5}$$

where it is important to note that, because of the nonlinearity in $\Phi(.)$ we have placed a restriction on β_p to be constant across the $\widehat{\mathbf{P}}_n$ and $\Delta\widehat{\mathbf{P}}_n$. Statistical discrimination, then, is indicated if the term $\beta_p \Delta\widehat{\mathbf{P}}_n$ is nonzero.

3.3 Evaluating Discrimination

Standard (classical) tests for discrimination might examine the statistical significance of the coefficients on the $\mathbf{x}_n$'s and perform a model comparison between (1) and (5). We will opt for a Bayesian environment in which we can assess the probability that discrimination is present in the sample. To accomplish this, we augment the rate equation with two vectors of model indicator dummies, γ and δ:

$$R_n = \beta_0 + \beta_p \left(\delta\widehat{\mathbf{P}}_n + (1-\delta)\,\widetilde{\mathbf{P}}_n\right) + \beta_z \mathbf{z}_n + \gamma \odot \beta_x \mathbf{x}_n + e_n,$$

where $\odot$ denotes the Hadamard product. The model indicators γ and δ are vectors of zeros and ones with dimensions $(\kappa_x \times 1)$ and $(\pi \times 1)$, respectively. Individual elements of γ will determine the extent of disparate treatment, disparate impact, or redlining in the rate. Because we restrict β_p to be the same in both the $\widehat{\mathbf{P}}_n$ and $\widetilde{\mathbf{P}}_n$ terms, δ can be thought of as a model selection variable that determines the presence of statistical discrimination.

3.4 Estimation

The rate equation (1), utilizes predicted performance and, therefore, suffers from a generated regressor problem (see Pagan, 1984). In a classical environment, one could estimate the probit model using, say, maximum likelihood and employ a bootstrap to estimate the standard errors (see Kilian, 1998). Instead, we opt to estimate the model in a Bayesian environment. We employ a set of relatively uninformative standard priors. The slope coefficients in both the rate equation and in the probit have mean zero normal priors; the variance of the innovations in the rate equation has an inverse Gamma prior. The priors for each of the model indicators are flat.

The posteriors used for inference are generated from the Gibbs sampler using two Metropolis-in-Gibbs steps. The Gibbs sampler is a Markov Chain Monte Carlo technique which iteratively draws each parameter from its conditional distribution. The collection of draws converges to the full set of parameters' joint posterior. Inference is performed on a subset of draws, some of which are discarded to allow for convergence.

Our algorithm is a three step procedure. In the first step, we draw the slope parameters of the probit. After allowing for convergence, for each draw of α, we compute two predicted performance measures, $\widehat{\mathbf{P}}_n$ and $\widetilde{\mathbf{P}}_n$, conditional on the draw of α. For each $\widehat{\mathbf{P}}_n$ and $\widetilde{\mathbf{P}}_n$ combination, we then iteratively draw 1000 samples of β, δ, and γ, burning the first 500 to account for convergence. The first step is repeated 500 times after convergence is achieved. We store β, δ, and γ draws every 10 draws, which yields 500 draws of α and 25,000 draws of β, δ, and γ, which are pooled. Note that the sampling algorithm described here accounts for the sampling uncertainty in α which would create the generated regressor problem in $\widehat{\mathbf{P}}_n$ and $\widetilde{\mathbf{P}}_n$. The final result is a set of posterior distributions for α and β and a set of model inclusion probabilities for each of the $\Delta\widehat{\mathbf{P}}_n$'s and $\mathbf{x}_n$'s.

Details of the sampling methods, including the specifications for the priors and the posterior draws, are included in the attached appendix.

4 Results

To implement the evaluation discussed in the previous section, we randomly divide the sample for each mortgage product in half. We use the first half to form the actuarial sample and estimate the probit model for two measures of loan performance: default within 2 years and prepayment within 2 years of closing.[5]

[5]We consider a loan in default if the LP variable MBA_STAT takes a value of 9, F, or R. We consider a loan prepaid if the loan leaves the database or has an MBA_STAT of 0 in a particular month and the MBA_STAT variable does not take a value of 6, 9, F, or R in the month before the loan leaves the database. To keep our model parsimonious, we do not construct loan performance measures for other horizons; see

For now we set $\delta = 1$ and leave the analysis of differentiating taste-based from statistical discrimination for future versions of this paper and we focus on the problem of identifying disparate treatment, disparate impact, and redlining.

Tables 7 and 8 present the results from the loan performance models using the actuarial sample. Table 7 present the results for the default measure, and table 8 presents the results for the prepayment measure. The tables present the medians of the posterior distributions of the coefficients. We indicate with an asterisk that 0 is *not* contained inside the corresponding 90 percent coverage interval. The results from the loan performance models indicate that standard measures of credit worthiness, such as FICO scores, loan-to-value ratios, and debt-to-income ratios are important determinants of both default and prepayment. Refinances are associated with lower default and higher prepayment. 30 year FRMs, 30 year ARMs, and 10 year FRMs are more likely to default in Florida than in California, while most mortgage products are less likely to be prepaid in Florida than in California. Loans for blacks and Hispanics are more likely to default in four of the eight mortgage product categories. Prepayment penalties on black and Hispanics appear to be associated with lower default rates, but seem to have a positive impact on the probability of prepayment for 2/28 ARMs and no apparent effect on other mortgage products. Higher tract income and higher share of tract minority population are associated with both lower default probability and higher prepayment probability.

Table 9 presents the estimation results in the underwriting sample. As before, the coefficients represent the medians of the posterior distribution and the asterisk indicates that 0 is not contained in the 90 percent coverage interval. However, the coefficients associated with the treatment variables $\mathbf{x}$ represent the medians of the posterior distributions, conditional on the mode of the corresponding inclusion variable γ, for cases in which the variable inclusion probability (the probability that the corresponding value of γ is equal to 1) exceeds 80 percent.

Demyanyk (2009) for evidence on the large proportion of subprime loans that terminate within two or three years of origination.

Table 7: Probit performance estimation. Default within 2 years

	Variable	2-28 ARM	3-27 ARM	30yr FRM	30yr ARM	10yr FRM	10yr ARM	5yr ARM	Other
	Constant	1.2551*	0.9636*	1.2140*	1.2601*	1.2362*	0.6653*	1.4518*	1.5754*
q	Loan-to-value	0.0616*	0.1424*	0.1533*	0.1560*	0.2138*	0.2207*	0.1325*	0.2417*
	Prepayment penalty	0.1818*	0.3809*	0.2458*	0.2190*	0.0956	0.2304*	0.3222*	0.2976*
	Debt-to-income	-0.0235*	-0.0196*	-0.0138*	0.0189*	0.0172	0.0057	-0.0008	0.0459*
	FICO score	-0.3547*	-0.4393*	-0.5093*	-0.4731*	-0.5004*	-0.4870*	-0.4589*	-0.5622*
	Private mortg. ins.	-0.0458	-0.0306	-0.1216*	-0.0261*	-0.0980	-0.4393*	-0.0031	-0.0265
	Loan amount	0.1459*	0.1982*	0.2724*	0.3172*	0.1374	0.1712*	0.2947*	0.1879*
	Full Doc	-0.2200*	-0.2679*	-0.1443*	-0.1662*	-0.4208*	-0.2759*	-0.1994*	-0.2367*
	Refinacing loan	-0.4690*	-0.3557*	-0.2014*	-0.2992*	-0.2578*	-0.2471*	-0.3987*	-0.4891*
	State=Florida	0.0494	0.0034	0.1843*	0.0880*	0.2188*	-0.0051	-0.0100	-0.1675*
	Borrower income	0.1105*	0.0033	-0.4022*	0.0552	0.1065	-0.0687	0.0335	0.1793
w	Black	0.1500	0.2279	0.2839*	0.2286*	0.0609	0.1057	0.2587*	0.2941*
	Hispanic	0.0927	0.0709	-0.0035	0.0802*	0.1951*	0.1936*	0.1635*	0.0804
	Prepay x black	-0.0080	-0.2805	-0.2368*	-0.1030	0.2149	0.1981*	-0.1533*	-0.1213
	Prepay x Hispanic	-0.1316	-0.1773	-0.1185*	-0.0542*	-0.1968	-0.0386	-0.0942*	-0.0233
	PMI x black	0.2569*	0.0582	-0.0340	-0.0486	-0.0462	0.0678	0.0226	0.0696
	PMI x Hispanic	0.0378	-0.1181	0.0568	-0.0278	-0.9076*	0.1928	0.0270	0.0246
	Tract income	-0.0245	-0.0209	-0.0519*	-0.0424*	-0.0765*	-0.0650*	-0.0512*	-0.0620*
	Tract minority	-0.0211*	0.0092	-0.0096*	-0.0150*	-0.0311*	-0.0142*	-0.0216*	-0.0295*

The coefficients represent the medians of the posterior distributions.
The asterisk indicates that 0 is not contained in the 90 percent coverage interval.
Tract income is equal to the census tract median family income relative to the HUD estimate of the metropolitan area's family income provided in the HMDA data.
Tract minority is the census tract percent of minority population from the 2000 census.

Table 8: Probit performance estimation. Prepayment within 2 years

	Variable	2-28 ARM	3-27 ARM	30yr FRM	30yr ARM	10yr FRM	10yr ARM	5yr ARM	Other
	Constant	0.7452*	1.4793*	1.3788*	0.9560*	1.6194*	0.1498	0.9001*	0.5232*
q	Loan-to-value	-0.0253*	-0.0214	0.0530*	-0.0527*	0.0015	0.0217*	-0.0334*	-0.0173*
	Prepayment penalty	-1.1506*	-0.5059*	-0.1763*	-0.4638*	-0.2454*	-0.2904*	-0.4948*	-0.2947*
	Debt-to-income	0.0177*	-0.0194*	0.0232*	0.0008	-0.0212*	-0.0173*	0.0137*	-0.0049
	FICO score	-0.0097	-0.1825*	-0.2646*	-0.0587*	-0.2865*	-0.0925*	-0.1245*	-0.1127*
	Private mortg. ins.	-0.0221	0.1351*	0.0989*	0.1199*	0.3502*	0.0702	0.1262*	0.0315
	Loan amount	-0.1540*	-0.1913*	-0.5321*	-0.1834*	-0.1588*	-0.0355	-0.2733*	0.0039
	Full Doc	-0.0233	-0.0882*	-0.0854*	0.0133*	-0.0843*	-0.1675*	-0.0648*	-0.1504*
	Refinancing loan	0.5527*	0.2766*	0.1123*	0.2389*	0.0448	0.0746*	0.4212*	0.3261*
	State=Florida	-0.1539*	-0.0750*	-0.2267*	-0.2357*	-0.0509	-0.2401*	-0.1289*	-0.1615*
	Borrower income	0.0478	0.1254*	-0.1239	-0.1351*	0.0974	0.1296	0.1182*	-0.2316*
w	Black	-0.3401*	0.0835	0.1032	-0.0630	0.1278	-0.0085	-0.0406	-0.0080
	Hispanic	-0.2369*	-0.0624	-0.0075	-0.0424*	0.0418	-0.0345	-0.0085	0.0571
	PPP x black	0.4260*	0.0866	-0.0867	0.0434	0.0262	0.0242	0.0584	-0.0202
	PPP x Hispanic	0.2075*	0.1089	0.0060	0.0158	0.0868	-0.0589	-0.0476*	-0.1549*
	PMI x black	-0.1194	-0.2325	-0.0316	-0.0139	-0.6196	-0.0733	-0.0346	0.0357
	PMI x Hispanic	0.0397	-0.1406	-0.0157	-0.0626*	-0.4981*	-0.0336	-0.0742*	-0.0602
	Tract income	0.0643*	0.0687*	-0.0070	0.0189*	0.0188	0.0209	0.0574*	0.0109
	Tract minority	0.0507*	0.0302*	0.0236*	0.0277*	0.0386*	0.0349*	0.0512*	0.0381*

The coefficients represent the medians of the posterior distributions.
The asterisk indicates that 0 is not contained in the 90 percent coverage interval.
Tract income is equal to the census tract median family income relative to the HUD estimate of the metropolitan area's family income provided in the HMDA data.
Tract minority is the census tract percent of minority population from the 2000 census.

Table 9: Rate Estimation

	Variable	2-28 ARM	3-27 ARM	30yr FRM	30yr ARM	10yr FRM	10yr ARM	5yr ARM	Other
	Constant	5.6566*	5.3611*	5.2537*	2.1197*	5.8899*	4.8991*	5.3230*	4.8815*
$\widehat{\mathbf{P}}$	Predicted default	4.9552*	5.8610*	5.0955*	11.2421*	3.9745*	3.7439*	4.3059*	4.7683*
	Predicted prepayment	2.2044*	1.5461*	3.1966*	5.6681*	0.7308*	2.3790*	2.1783*	3.3832*
$\mathbf{z}$	Prepay penalty	-0.1425	0.0680	0.1381*	0.4140*	-0.0268	0.2145*	0.0495*	0.0100
	Private mortg. ins.	0.2703*	0.1852	-0.0194	0.3341*	-0.0649	0.2417*	0.1053*	0.2017*
	Loan amount	-0.1134*	-0.1932*	0.0549	-1.4858*	0.0605	-0.0275	-0.1691*	-0.3703*
	State=Florida	0.4817*	0.4436*	0.4716*	0.8523*	0.1293*	0.3147*	0.4889*	0.9763*
$\mathbf{x}$	Black				0.1571*			0.1366*	0.3701*
	Hispanic				0.0917*	0.0547*	0.0701*	0.1343*	0.2344*
	Prepay x black						0.2258*		-0.3121*
	Prepay x Hispanic							-0.0930*	-0.1506*
	PMI x black				-0.2729*				
	PMI x Hispanic				-0.1372*				
	Tract income	-0.0578*		-0.0269*	-0.0610*	-0.0362*		-0.0462*	-0.0668*
	Tract minority						0.0121*	0.0071*	

The coefficients of the **z** variables represent the medians of the posterior distributions.
The coefficients of the **x** variables represent the medians of the posterior distributions conditional on the modal value of the corresponding γ for cases in which the inclusion probability $\Pr(\gamma = 1)$ exceeds 80 percent.
The asterisks indicates that 0 is not contained in the 90 percent coverage interval.

The results from table 9 indicate that both measures of forecasted performance have a positive impact on rate determination. Prepayment penalties and private mortgage insurance requirements also increase rates in about half of the mortgage product categories. Higher loan amounts reduce interest rates, and loans in Florida exhibit higher interest rates than in California.

As in Haughwout et al. (2009) we find no evidence of discrimination in 2/28 ARMs. However, lower-income neighborhoods face higher interest rates for this mortgage product, which suggests redlining. Disparate treatment discrimination does appear to be present in other mortgage products. Specifically, race indicators are associated with higher interest rates in 30 year ARMs, 10 year FRMs, 10 year ARMs, 5 year ARMs, and in the remainder category. The Hispanic indicator has a positive impact on all of these categories, while the black indicator has a positive effect only on 30 year ARMs, 5 year ARMs, and in the remainder category. Hispanics appear to face prepayment penalties as a requirement for obtaining lower interest rates in 5 year ARMs and in the remainder category. The interaction of the indicator for blacks and prepayment penalties has a positive effect on rates in 10 year ARMs and a negative effect in the remainder category. Purchase of private mortgage insurance among black and Hispanics also lowers interest rates in 30 year FRMs.

Redlining, as indicated from lower tract income associated with higher interest rates, appears to be present not only in 2/28s, but also in 30 year FRMs, 30 year ARMs, 10 year FRMs, 5 year ARMs, and the remainder category. Furthermore, a higher share of minorities also leads to higher interest rates in 10 year ARMs and in 5 year ARMs.

5 Conclusions

In this paper we examined the effect of race and ethnicity on the pricing of subprime mortgages in California and Florida during 2005. We estimated a reduced-form model of mortgage rate determination in which the lender takes into account the predicted loan performance when making the rate-setting decision. We assessed the effect of race and ethnicity, as well as the effect of neighborhood characteristics, both in the loan performance evaluation and in the lender's rate decision.

In contrast with previous studies of the subprime market we find evidence of adverse pricing for black and Hispanic borrowers in many of the mortgage products we considered. These effects are substantial and lead to rate increases ranging from 5 to 35 basis points. We also find an adverse pricing effect in lower income neighborhoods. Finally, we find that for minorities, the purchase of private mortgage insurance and prepayment penalty fees seem to be associated with obtaining lower interest rates.

References

Demyanyk, Yuliya, 2009. "Quick Exits of Subprime Mortgages". *Federal Reserve Bank of St. Louis Review*, March-April, 79-94.

Haughwout, Andrew; Mayer, Christopher; and Tracy, Joseph, 2009. "Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing". Federal Reserve Bank of New York Staff Report no. 368.

Holmes, Chris C. and Held, Leonhard, 2006. "Bayesian auxiliary variable models for binary and multinomial regression". *Bayesian Analysis* 1, 145-168.

Kilian, Lutz, 1998. "Small-Sample Confidence Intervals for Impulse Response Functions". *Review of Economics and Statistics* 80, 218–230.

Ladd, Helen F., 1998. "Evidence of Discrimination in Mortgage Lending". *Journal of Economic Perspectives* 12:2, 41-62.

Mayer, Christopher J. and Pence, Karen, 2008. "Subprime Mortgages: What, Where, and to Whom?" NBER Working Paper 14083.

Munnell, Alicia H.; Browne, Lynn E.; McEneaney, James; and Tootell, Geoffrey M.B., 1996. "Mortgage Lending in Boston: Interpreting HMDA Data". *American Economic Review* 86:1, 25-53.

Pagan, Adrian, 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors". *International Economic Review* 25, 221-247.

Pope, Devin G. and Sydnor, Justin R., 2008. "What's in a Picture? Evidence of Discrimination from Prosper.com". Manuscript, Case Western Reserve University.

Ravina, Enrichetta, 2008. "Love and Loans: The Effect of Beauty and Personal Characteristics in Credit Markets". Manuscript, Columbia University.

Reid, Carolina and Laderman, Elizabeth, 2009. "The Untold Costs of Subprime Lending: Examining the Links among Higher-Priced Lending, Foreclosures and Race in California". Manuscript, Federal Reserve Bank of San Francisco.

Ross, Stephen L. and Tootell, Geoffrey M.B., 2004. "Redlining, the Community Reinvestment Act, and Private Mortgage Insurance". *Journal of Urban Economics* 55, 278-297.

Ross, Stephen L. and Yinger, John, 2002. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement.* MIT Press: Cambridge, Massachusetts.

Tanner Martin A. and Wong, Wing Hung, 1987. "The Calculation of Posterior Distributions by Data Augmentation". *Journal of the American Statistical Association*, 82, 528-540.

Troughton, Paul T. and Godsill, Simon J., 1997 "A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves". Technical Report CUED/F-INFENG/TR.304, Cambridge University Engineering Department.

Woodward, Susan E., 2008. *A Study of Closing Costs for FHA Mortgages.* U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

A Estimation Details

This appendix describes the Bayesian methods used to estimate the model in Section 3. The model is estimated with an iterative technique – the Gibbs sampler – which requires a prior. For the slope parameters in both (5) and (2), we assume a normal prior. The innovation variance of the rate equation has an inverse Gamma prior. Each of the model indicators has a flat prior. The hyper-parameters for the prior distributions are shown in table 10.

Estimation of the parameters of (2) can be accomplished by data augmentation (Tanner and Wong, 1987). Define a latent variable y_{im} which has mean $\alpha_{i0} + \alpha_{iq}\mathbf{q}_m + \alpha_{iw}\mathbf{w}_m$, unit variance, and is restricted such that $y_{im} > 0$ iff $P_{im} = 1$. Then, conditional on α_i, $y_i = \{y_{im}\}_{m=1}^{M}$ can be drawn independently from truncated normal distributions. Let $\mathbf{q} = (q_1, ..., q_M)'$ and $\mathbf{w} = (w_1, ..., w_M)'$. Then, conditional on the drawn y_{im}, we draw α_i from a normal posterior:

$$\alpha_i | y_i \sim N(\mathbf{a}_i, \mathbf{A}_i),$$

Table 10: Priors for Estimation

Parameter	Prior Distribution	Hyperparameters
α_i	$N(\mathbf{a}_0, \mathbf{A}_0)$	$\mathbf{a}_0 = \mathbf{0}_{1+\kappa_q+\kappa_w}$; $\mathbf{A}_0 = \mathbf{I}_{1+\kappa_q+\kappa_w}$
β_{-p}	$N(\mathbf{b}_0, \mathbf{B}_0)$	$\mathbf{b}_0 = \mathbf{0}_{1+\kappa_x+\kappa_z}$; $\mathbf{B}_0 = \mathbf{I}_{1+\kappa_x+\kappa_z}$
β_p	$N(\mathbf{d}_0, \mathbf{D}_0)$	$\mathbf{d}_0 = \mathbf{0}_\pi$; $\mathbf{D}_0 = \mathbf{I}_\pi$
σ^{-2}	$\Gamma\left(\frac{\nu_0}{2}, \frac{\Upsilon_0}{2}\right)$	$\nu_0 = 6$; $\Upsilon_0 = 0.01$

where $\mathbf{a}_i = \left(\mathbf{A}_0^{-1} + \mathbf{X}_i'\mathbf{X}_i\right)^{-1}$, $\mathbf{m}_i = \mathbf{M}_i\left(\mathbf{M}_0^{-1}\mathbf{m}_0 + \mathbf{X}_i'\mathbf{y}_i\right)$, and $\mathbf{y}_i = (y_{i1}, ..., y_{iM})'$, and $\mathbf{X}_i = (\mathbf{1}_M, \mathbf{q}, \mathbf{w})$. After a suitable number of draws are discarded to obtain convergence, we utilize the draws of the α_i to generate predictions for performance of the N loans to be used for underwriting. For each draw, we compute $\widehat{\mathbf{P}}_n$ and $\widetilde{\mathbf{P}}_n$ from (3) and (4), respectively.

For each (post convergence) draw of $\widehat{\mathbf{P}}_n$, we sample 1000 draws from the posterior distributions of the model parameters β_{-p}, β_p, γ, δ, and σ^2. Conditional on δ and σ^2, the model inclusion parameters, γ, and the vector of slopes (excluding β_p), β_{-p}, can be drawn jointly from a reversible jump Metropolis Hastings in Gibbs step (see Troughton and Godsill, 1997, and Holmes and Held, 2006).[6] The joint move uses a proposal density of the form:

$$q\left(\gamma^*, \beta_{-p}^*; \gamma, \beta_{-p}\right) = p\left(\beta^* | \gamma^*, \beta_{-p}\right) q\left(\gamma^* | \gamma\right),$$

which means we draw the candidate γ^* first and then, conditional on γ^*, we draw β_{-p}^*. The candidate γ^* is generated by drawing a random index from a discrete uniform distribution. The element corresponding to the drawn index is switched – one to zero, zero to one. Then, conditional on γ^*, the prior for β_{-p} is

$$\beta_{-p}^* \sim N\left(\mathbf{b}_0^*, \mathbf{B}_0^* | \gamma^*\right),$$

where $\mathbf{b}_0^*$ and $\mathbf{B}_0^*$ are the hyperparameters corresponding to the candidate covariate set. The candidate β^* is drawn from

$$\beta_{-p} \sim N\left(\mathbf{b}^*, \mathbf{B}^* | \gamma^*\right),$$

with parameters:

$$\mathbf{b}^* = V\left(\mathbf{B}_0^{*-1}\mathbf{b}_0^* + \sigma^{-2}\zeta'\widetilde{\mathbf{R}}\right)$$

[6] Turning elements of the indicator γ on and off changes the model dimension. The resulting variation in the model dimension across Gibbs iterations makes joint sampling more efficient.

and

$$\mathbf{B}^{*}=\left(\mathbf{B}_{0}^{*-1}+\sigma^{-2}\zeta'\zeta\right)^{-1},$$

where $\mathbf{R}=\left(R_{1}-\beta_{p}\left(\delta\widehat{\mathbf{P}}_{1}-(1-\delta)\,\widetilde{\mathbf{P}}_{1}\right),...,R_{N}-\beta_{p}\left(\delta\widehat{\mathbf{P}}_{N}-(1-\delta)\,\widetilde{\mathbf{P}}_{N}\right)\right)'$, $\zeta_{n}=(1,\mathbf{z}_{n}',\mathbf{x}_{n}')'$, and $\zeta=(\zeta_{1},...,\zeta_{N})$. We accept the joint draw $\left[\gamma^{*},\beta_{-p}^{*}\right]$ with probability

$$A=\min\left\{1,\frac{|\mathbf{B}_{0}|^{1/2}}{|\mathbf{B}_{0}^{*}|^{1/2}}\frac{|\mathbf{B}^{*}|^{1/2}}{|\mathbf{B}|^{1/2}}\frac{\exp\left(\frac{1}{2}\mathbf{b}^{*}\mathbf{B}^{*-1}\mathbf{b}^{*}\right)}{\exp\left(\frac{1}{2}\mathbf{b}\mathbf{B}^{-1}\mathbf{b}\right)}\right\},$$

where $\phi\left(.\right)$ is the multivariate normal pdf and the unstarred $\mathbf{b}$, $\mathbf{B}$, and $\mathbf{B}_{0}$ correspond to the hyperparameters computed conditional on the last (accepted) iteration of γ.

Next, we draw the joint pair (δ,β_{p}) by again selecting a candidate δ^{*} and drawing β_{p}^{*} from a normal proposal, conditional on δ. The proposals for δ and β_{p} – as well as the acceptance probability – have forms similar to those expressed above. For brevity, we omit the formalities.

The final step in the Gibbs loop is the draw of σ^{2} conditional on β_{-p}, β_{p}, γ, δ, and the data. Given the prior, the innovation variance can be drawn from the inverse gamma posterior

$$\sigma^{-2}|\gamma,\delta,\beta,\mathbf{R}\sim\mathbf{\Gamma}\left(\frac{\nu_{0}+N}{2},\frac{\Upsilon_{0}+\mathbf{e}\prime\mathbf{e}}{2}\right),$$

where $\mathbf{e}=\mathbf{R}-\beta\zeta$ and $\zeta=\left(\mathbf{1}_{N},\delta\widehat{\mathbf{P}}_{N}-(1-\delta)\,\widetilde{\mathbf{P}}_{N},\mathbf{z}_{N}',\mathbf{x}_{N}'\right)'$.